



HUMAN-COMPUTER INTERACTION AND INFORMATION MANAGEMENT RESEARCH NEEDS

Human-Computer Interaction and Information Management
(HCI&IM) Coordinating Group

Interagency Working Group on
Information Technology Research and Development
(IWG/ITR&D)

October 2003

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE OCT 2003		2. REPORT TYPE		3. DATES COVERED 00-00-2003 to 00-00-2003	
4. TITLE AND SUBTITLE Human-Computer Interaction and Information Management Research Needs				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) National Coordination Office ,Natworking and Information Technology Research and Development,4201 Wilson Boulevard Suite II-405,Arlington,VA,22230				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 50	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

This page intentionally left blank.

TABLE OF CONTENTS

REPORT OVERVIEW	1
1. INTRODUCTION AND PRIMER.....	3
1.1 Why Do Humans Use Computing Systems?	4
1.2 What Do Humans Use Computing Systems For?.....	6
1.3 How Do Humans and Computing Systems Use Data and Information?	7
1.3.1 How Do Humans Use Data and Information?	11
1.3.2 How Do Computing Systems Use Data and Information?	11
1.3.3 How Can Humans and Computing Systems Best Work Together?	12
2. HCI&IM RESEARCH NEEDS.....	15
2.1 Information Creation, Organization, Access, and Use	15
2.2 Managing Information as an Asset	16
2.3 Human-Computer Interaction and Interaction Devices	17
2.4 Evaluation Methods and Metrics	18
3. HCI&IM CG GOVERNMENT WORKSHOP - OCTOBER 22, 2001	21
3.1 DARPA	21
3.2 EPA	22
3.3 DOE/SC	23
3.4 NOAA	24
3.5 NIH	25
3.6 NSF	27
3.7 NASA	28
3.8 NIST	30
3.9 GSA	31
3.10 AHRQ	33
4. OTHER GOVERNMENT PROGRAMS	35
4.1 NSF's Digital Libraries and Digital Government Programs and their Joint Workshop with the Library of Congress on Preservation.....	35
4.2 Air Force Research Laboratory Human Effectiveness Directorate	36
4.3 Department of Energy/Office of Scientific and Technical Information (DOE/OSTI) and CENDI.....	37
APPENDIX 1: WORKSHOP ATTENDEES AND CONTRIBUTORS TO THIS REPORT.....	39
APPENDIX 2: ACRONYMS	43

TABLE OF FIGURES

Three representations of ocean surface speed around Antarctica:

Figure 1.	Ocean surface speed data stored as bits in a computer	8
Figure 2.	The ocean surface speed data with its metadata	9
Figure 3.	Images that use the same ocean surface speed data and its metadata to show differences in ocean speed over time and space	10

REPORT OVERVIEW

In a visionary future, Human-Computer Interaction (HCI) and Information Management (IM) have the potential to enable humans to better manage their lives through the use of synergistic human-computer empowered information systems.

This report identifies and illustrates the problems that underlie HCI&IM R&D and provides a context and snapshot of Government approaches to HCI&IM R&D for the development and use of integrated on-line information technologies to achieve benefits such as:

- Changing the way scientific research is conducted
- Expanding the science and engineering knowledge base
- Enabling a more knowledgeable, capable, and productive workforce

The report is built on a one-day multi-agency workshop held on October 22, 2001, that documented ongoing Government efforts and identified programmatic gaps where solutions are critical to meet agency mission goals and objectives.

Chapter 1 defines concepts and the vision for HCI&IM research needs, addresses several assumptions about the differences between the HCI&IM concepts including data and information, and provides a context for the research needed to address fundamental HCI&IM problems.

Chapter 2 identifies and explains four critical research areas that form the basis of the multi-agency HCI&IM research agenda and that characterize the scope of HCI&IM R&D. These four areas were derived from the findings and conclusions of workshop materials and presentations and from extensive subsequent discussions within the HCI&IM Coordinating Group (CG). They are:

- Information Creation, Organization, Access, and Use
- Managing Information as an Asset
- Human-Computer Interaction and Interaction Devices
- Evaluation Methods and Metrics

Chapter 3 provides an overview of FY 2001 Federal government agency programs and needs in HCI&IM R&D and suggests new areas to be used in planning future HCI&IM R&D investments. Ten Federal agencies that are members of the HCI&IM CG participated in the workshop. They are: AHRQ, DARPA, DOE/SC, EPA, GSA, NASA, NIH, NIST, NOAA, and NSF. (Acronyms are defined in Appendix 2.)

Chapter 4 adds descriptions of other Federal activities that are contributing to HCI&IM R&D.

This page intentionally left blank.

1. Introduction and Primer¹

Humans use information to address their desire and need to learn, communicate, explain, and make decisions. Information is stored in computing systems as “bits” that are either zero or one and collections of bits such as the eight-bit bytes used to encode up to 256 distinct letters, digits, and characters. A computing system’s bits and bytes are converted into spoken or written words, symbols, or images, for humans to understand. Humans define and develop both data and information and their use by computing systems so that they contribute to achieving human goals.

The purpose of the HCI&IM CG is to coordinate research and development of technologies for mapping human knowledge into computing systems and back, all for the purpose of human understanding, analysis, and use. The scope of HCI&IM R&D is the data, information, and associated processes used to meet the mission goals of the HCI&IM CG agencies. The scope also includes making the data, information, and output of those processes usable by and useful to humans with diverse requirements and capabilities.

Information management and use are among the greatest challenges facing our country. Computer-based information systems are essential in maintaining national and homeland security, infrastructure protection, science, health care, and managing resources such as the power grid. These information systems must operate continuously, provide instantaneous access to up-to-date information during emergencies, and evolve to meet the Nation’s ever-expanding information needs. How to meet these challenges set the starting point for developing this report to inform a prioritized Federal HCI&IM research agenda.

The HCI&IM R&D vision is to make possible the following benefits:

- Information that is available everywhere, at any time, and to everyone regardless of their abilities
- Broadened human use of this information through the ability to:
 - Access information in different contexts – examples include:
 - Having the same information presented differently for audiences as varied as school children and researchers
 - Providing the same content to different devices
 - Interact with this information using a variety of devices
 - Meet needs that can vary from access to manipulation, analysis, and control
- Comprehensive management of this vast information environment

Consistent with this vision, a recent National Research Council report entitled *IT Roadmap to a Geospatial Future*² explains the need for HCI&IM R&D. It stresses the critical role HCI&IM plays in emergency response to natural and man-made disasters.

¹ Two books were consulted in preparing this chapter: (1) C. Cherry, *On Human Communication*, MIT Press, 1957 and 1966, and (2) C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949 and 1998.

The nature of the R&D addressing these problems can be illustrated best by answering the following fundamental questions:

- *Why do humans use computing systems?*
- *What do humans use computing systems for?*
- *How do humans and computing systems use data and information?*
 - *How do humans use data and information?*
 - *How do computing systems use data and information?*
 - *How can humans and computing systems³ work synergistically and for what purposes?*

For this report, data and information are distinguished as follows:

- *Data* are raw collected measures, facts, or observations.
- In computer-based information systems, *information* includes both data and metadata that label the data. Metadata can include descriptions of the data, how they were collected, and numerous other features.

Computing systems process data and information differently.

- Examples of how computers process data are:
 - Using them in numerical calculations
 - Graphing them in x-y plots
 - Examining them to determine if patterns in different data sets match
- Information, on the other hand, can be processed by inference to derive relationships that are then validated through examining data.

For instance, the content of a cell in a database is a datum (singular of data). It may be a fact such as a person's name or address. When the cells in the database are labeled (name, address, etc.), their form defined (text, numbers, etc.), and the relationships among the cells described, then the database as a whole is information.

1.1 Why Do Humans Use Computing Systems?

For millions of years, humans have used tools to ease the tasks they need to perform in order to survive. From historical tools such as chiseled-rock spear points to tools of the 21st century, humans have used their innovative talents and their enriched understanding of science to create technologies and tools to support their needs. Computing systems are the latest and arguably

² National Academies Press, Washington, D.C., 2003.

³ The terms “computers” and “computing systems” are used interchangeably throughout this document. These terms encompass individual computational nodes, memory, storage, visualization other special purpose hardware, and the software that operates them. They include desktops, laptops, workstations, computational grids, supercomputers, hand-held personal digital assistants, networked sensors and actuators, and low-power computers on satellites.

most complex tools that humans have ever created. Their use continues to evolve and grow as their speed and capability increase.

From their earliest days, computers were used for scientific, engineering, and cryptographic computations. Early human users of computers both wrote and read the bits that computers understood. With the advent of programming languages and translators such as assemblers and compilers, humans used characters (numbers, letters, and punctuation) and words to write instructions in ways that humans understood, and which translators turned into computer-readable commands. Conversely, computers converted the bits that comprised the results of computations into numbers, words, and sentences that humans understood. Humans viewed the content created by computers as reflecting that computers could think, which greatly expanded the potential roles computers could play. Computers could not just compute, but could also produce sentences that people could understand.

Understanding what a computer is doing in human terms rather than in computer terms was an early step in human-computer interaction. As humans observed what computers could do, they adapted the computer's capabilities to satisfy their needs and desires. Today humans use computers to communicate, interact, and share in diverse ways:

- To write and talk to each other
- To exchange artifacts of personal interest such as photos, music, and videos
- To exchange artifacts used in the workplace such as text files, drawings, and visualizations

Ideas about what computers could do for humans grew as single stand-alone computers that performed only one set of calculations using data on external media such as tapes evolved into our multi-functional 21st century systems. More ideas came as networks of computing systems were built to transmit bits almost instantaneously around the world for other computers and software – and people – to use. In a seemingly never-ending cycle, new computer-enabled capabilities, devices, and environments are envisioned, developed, and deployed for use by ever-larger, more diverse, and more demanding populations. These human activities have greatly expanded computer usage.

Humans use computers to meet their needs and desires as long as they can easily do so. A user will not invest in learning to use a computer when such learning is disproportionate to meeting those needs and desires. User communities assess ease-of-use differently, and the great variability in human needs and capabilities means that ease-of-use has many aspects. For example, astrophysicists will invest extraordinary effort to use fast new (and difficult-to-use) computing systems to conduct their research because they have no viable alternative. Each science and engineering community has its own assumptions, data, information, vocabulary, computing needs, and ways to communicate, interact, and share. Each community has developed its own computing system capabilities. Building easily traversed bridges among such systems and across nations, languages, and cultures, is a demanding task.

1.2 What Do Humans Use Computing Systems For?

In little over 50 years, computing systems have become part of virtually every human activity. The following are key areas of interest to the Federal government in which computing systems are used:

- National defense and national security
- Homeland security
- Infrastructure operation and protection
- The workplace
- Education and training
- Financial, personnel, and clientele records management
- Health care
- Manufacturing
- Research and development (R&D) in science, engineering, and technology

R&D communities have developed computing system capabilities to meet their needs, such as:

- Scientific calculations that implement mathematical formulas and increasingly produce huge computer-based data sets
- Computer-based sensors that collect huge data sets
- Software programs that search for patterns in such data sets
- Computer-generated visualizations (images or videos created by controlled mapping of interpreted data bits) for human pattern recognition and for studying multiple data sets
- Computer-based information for use by other researchers

The broader population (which includes R&D communities) has found it beneficial to adapt these systems for new and different capabilities, such as:

- Moving more information on line by creating digital content and by converting material from non-digital form to digital form
- Using general-purpose computing systems to manage the telecommunications infrastructure or control industrial processes, for example
- Using special-purpose computing systems and computing systems that have special physical features, such as the small purpose-built computers in today's automobiles that communicate with computers in satellites to provide location information; the car's computers present this information in different forms to the car and to humans
- Using global-scale computer-based networks to access computing systems that contain the data and information and provide the services they wish to use; the Internet and the Web are examples
- Interacting with computing systems and each other either simultaneously or by connecting to networks and computing systems at different times; an example is school children around the world communicating with one other, contributing to on-going scientific data collection, and "seeing" natural phenomena such as California's Monterey Bay during a scientific dive

Through advances in computing and human interaction technologies and ease of use, human use of the on-line world will continue to grow.

1.3 How Do Humans and Computing Systems Use Data and Information?

Humans and computers use data and information differently. These differences have two aspects: the form of the data or information and the processing applied to the data or information. Computer-based content has to be presented to humans in forms that are different from internal computer form, and humans process that information in different ways than do computers. By understanding the functional capabilities and limitations of computing systems and of humans, we can create an on-line world that best serves human needs.

An example may be helpful. This example is of ocean surface speed around Antarctica, and has three parts:

- Figure 1 shows how the ocean surface speed data are stored as bits in a computer.
- Figures 2 and 3 show the same data in forms that are more easily understood by humans:
 - Figure 2 shows the data with its metadata, which is also stored on a computer. The bits have been grouped and mapped into letters, words, numbers, and characters that describe the content in a way that humans can understand and use, and that humans can write and a computer can then translate back into bits.
 - Figure 3 is an alternative representation of the same bits, in this case a map that is part of an animation that can show differences in ocean speed over time and space. Humans are able to see patterns in such representations.

This example illustrates the breadth of the problems that HCI&IM R&D is addressing:

- To be able to cycle from bits to various humanly comprehensible representations of data and information, and in some cases to cycle back to bits
- To enable humans to interact with the content in its various forms in order to obtain additional information (from the bits) and to manipulate representations to better understand content

OCEAN SURFACE SPEED DATA STORED AS BITS IN A COMPUTER

```
0010101001011110101110110101010100111000111111010101011111101100110101001000  
011111111010010  
1111101001010100101101100110110100111111101110111101100110101001000011100101  
101100110110100  
110101010011010010101001010100110100110101010011010100101010001010100110100110  
100101010011010  
011 ....
```

Figure 1. Small subset of a gigabyte (10^9 bytes) file containing a representation of ocean surface speed. The file was produced by the Hybrid Isopycnal Model (HIM)⁴ at NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) in Princeton, New Jersey.

⁴ http://www.gfdl.noaa.gov/~lat/webpages/om/om_webpage.html

OCEAN SURFACE SPEED DATA WITH ITS METADATA

```
netcdf D.360.112.23 {
dimensions:
lath = 112 ;
latq = 113 ;

lonh = 360 ;
longq = 360 ;
Layer = 23 ;
variables:
double lath(lath) ;
    lath:long_name = "Latitude" ;
    lath:units = "degrees" ;
double latq(latq) ;
    latq:long_name = "Latitude" ;
    latq:units = "degrees" ;
double lonh(lonh) ;
    lonh:long_name = "Longitude" ;
    lonh:units = "degrees" ;
    lonh:modulo = "" ;
double longq(longq) ;
    longq:long_name = "Longitude" ;
    longq:units = "degrees" ;
    longq:modulo = "" ;
int Layer(Layer) ;
    Layer:units = "Layer" ;
double D(lath, lonh) ;
    D:long_name = "Basin Depth" ;
    D:units = "meter" ;
double g(Layer) ;
    g:long_name = "Reduced gravity" ;
    g:units = "meter second-2" ;
double R(Layer) ;
    R:long_name = "Target Potential Density" ;
    R:units = "kilogram meter-3" ;
double f(latq, longq) ;
    f:long_name = "Coriolis Parameter" ;
    f:units = "second-1" ;
lath = -73.7419122788742, -73.4595902819108, -73.172505031753,
-72.8805834944081, -72.5837519036926, -72.2819357756051,
-71.9750599242025, -71.6630484790547, -71.3458249043564,
-71.0233120197763, -70.6954320231271, -70.3621065149412,
-70.0232565250407, -69.6788025411921, -69.3286645399374,
-68.9727620196979, -68.6110140362472, -68.2433392406516,
-67.8696559197802, -67.4898820394868, -67.1039352905673,
-66.7117331375991, -66.3131928707682, -65.9082316607918,
-65.4967666170443, -65.0787148489947, -64.6539935310628,
-64.2225199710027, -63.784211681918, -63.3389864580144, ... }
```

Figure 2. A different representation of the same ocean surface speed data. This representation includes metadata that describe the data, including how many bits are used to represent a number. The metadata are then used to translate the bits in Figure 1 into numbers. Humans can scan these numbers to detect trends or discover points of significant shifts in the numerical values.

IMAGES OF OCEAN SURFACE SPEED DATA AND ITS METADATA

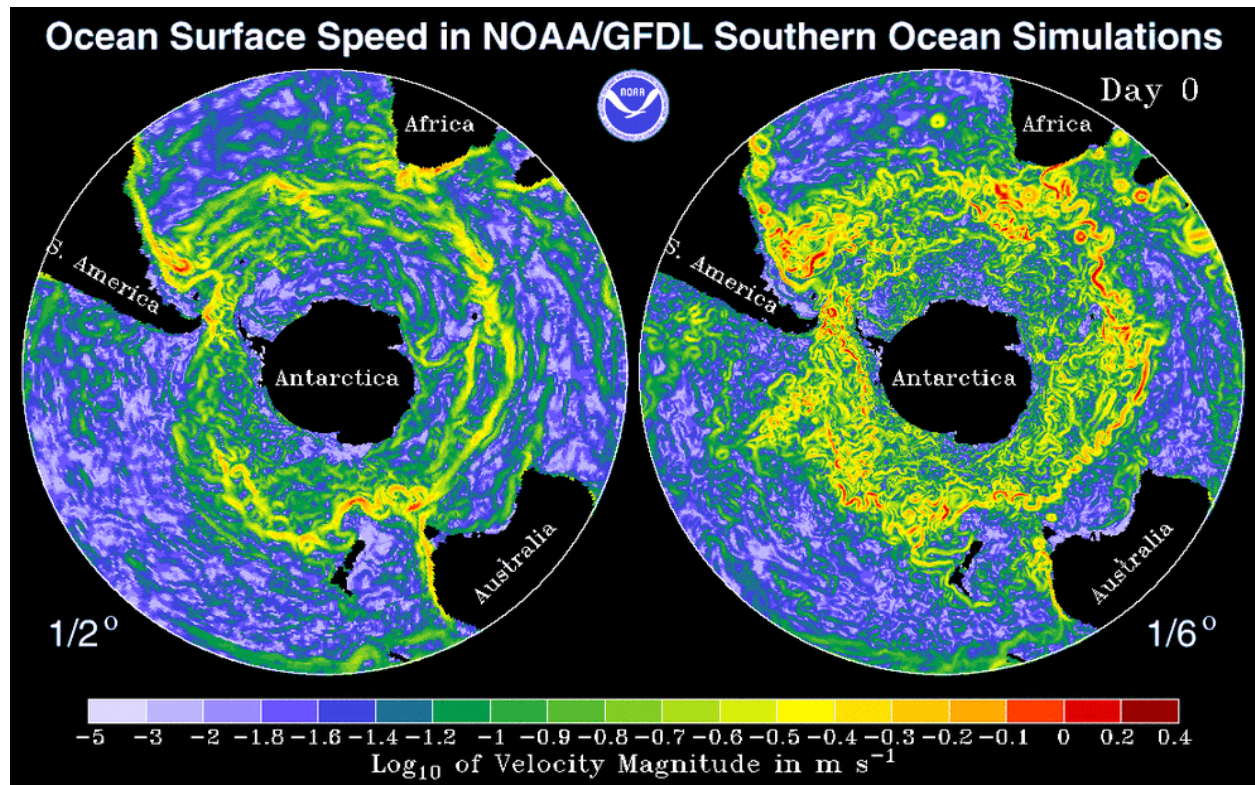


Figure 3. This third representation has two color-coded images of the same ocean surface speed data around Antarctica; the image on the right has greater detail. Scientists can watch an animation that includes these images in order to detect variations in speed for the same area.

1.3.1 How Do Humans Use Data and Information?

Humans use data and information to:

- Understand and learn about the world from direct observation
- Understand and learn about the world from artifacts (for example, off-line artifacts such as books and sculptures, and data and information – which include digital representations of books and sculptures – in on-line artifacts)
- Create new information
- Make decisions
- Control processes (such as the operation of a nuclear power plant)
- Communicate with other people
- Communicate with computing systems
- Share what they have learned and created with others
- Explain, inform, and teach

Humans process information through their senses, by induction (deriving general principles from particular facts or instances), and by deduction (the process of reasoning in which a conclusion follows necessarily from the stated premises; inference by reasoning from the general to the specific).⁵ Neural or nerve-based processes are the basis of all human processes, including processing information. Neural processes are continuous, dynamic (the same input can result in different responses), and operate in parallel. Humans excel at induction. For example, they can instantaneously detect a pattern in a visual image. Humans have difficulty with deduction, using statistical results in consistent ways, and in using measures of uncertainty when making decisions. Humans validate their views of the world through experimentation, theory, and, more recently, computer-based modeling and simulation. Only humans can validate information resulting from these processes.

1.3.2 How Do Computing Systems Use Data and Information?

The operations that a computer performs, whether numerical or logical, are specified by humans. The input to and the output from these operations are bits (sometimes organized into bytes) that have either been created digitally or converted from other forms. The input and output can be data or information, which a computer knows where to find because it is told the location explicitly (by a computer program that was created by humans). Ideally, the correctness of these operations and locations are verified before execution.

The tasks that computers can do better (meaning for example faster, more accurately, or at larger scale) than humans include:

⁵ The American Heritage Dictionary of the English Language, Fourth Edition, Houghton Mifflin Company, New York, 2000.

- Perform numerical calculations
- Use mathematics to model and simulate real-world phenomena
- Compare data sets and find statistical patterns
- Use mathematical logic to reason deductively about information
- Make decisions (based on rules established in computer processes set up by humans)
- Connect with other computers. Reasons for doing so include to:
 - Collect and find data and information (results of a google search of the Web are an example) and download or send what is found
 - Collect data from computer-based sensors placed in the real world
 - Send instructions to computer-based actuators placed in the real world

However, computers are currently less able than humans to:

- Create information
- Create new ways to communicate effectively with humans
- Validate that their output is true (for example, that it accurately represents the real world)

When humans allow computers to send instructions to actuators, those instructions can affect the real world, for example by shutting down a nuclear power plant. Computing systems need to demonstrate more stable and more reliable performance for humans to trust them to perform life-critical operations. Even then, in order that informed decisions can be made, (1) humans will sometimes need to monitor and be able to override computer operations and (2) the computers will need to be programmed to provide information and interact with their human managers.

1.3.3 How Can Humans and Computing Systems Best Work Together?

A widely held vision considers computing technologies to be essential tools for many human activities. The previous sections compared and contrasted the capabilities and limitations of human systems and computing systems. Current research focuses on identifying when and how to mix the best of human capabilities with current and future computing capabilities in order to maximize the use of computers to serve and enhance humans. A synergistic approach can provide richer capabilities than could be possible for either humans or computers alone. This is a complex problem, and it may be useful to consider scenarios:

Air Traffic Management

Some of today's aircraft have computer-based capabilities that would enable them to take off, fly, and land safely, independent of pilot control. These capabilities are only partially employed today, and are overseen by humans, because the computer-based systems cannot determine the state of their external operating environment as well as humans can. In this and similar situations, humans will need to determine:

- When to safely deploy an independent computer-based system
- How best to maintain human involvement in that system's operation

Human Genome Science

This relatively new and promising field provides opportunities to explore hybrid human/computer systems that capitalize on human perceptual, visual, and analytical processing skills and on computing systems' skill in pattern recognition. By learning from successive working systems and their use, we can leverage our understanding of the computing system's performance capabilities and human performance needs to better map computer results into human terms to improve the creation and validation of new genomic information.

Such scenarios can help us explore the range of human abilities, needs, and desires. Along with lessons learned about current computing systems' capabilities and new research in computing and in HCI&IM, such scenarios can help us develop rich computing systems tools with which humans can interact as they advance their understanding of and interactions with each other and the world around them.

This page intentionally left blank.

2. HCI&IM Research Needs

This chapter details HCI&IM R&D needs that are broadly shared by the HCI&IM CG agencies that were identified in their workshop presentations (chapter 3) and subsequent fact finding (chapter 4) and discussion. They are organized into four areas:

- Information Creation, Organization, Access, and Use
- Managing Information as an Asset
- Human-Computer Interaction and Interaction Devices
- Evaluation Methods and Metrics

Successes will help enable robust solutions to meet agency mission requirements and national needs.

2.1 Information Creation, Organization, Access, and Use

Many workshop presentations included descriptions of agency needs for very large data and information repositories that can be used in distributed ways by a variety of on-line communities. The repositories are primarily to be used to support scientific investigations, but some agencies also need to support data collection and creation, decision making, and some provide information to wider populations for other purposes, such as education. To better support all of these uses, the agencies are both conducting R&D toward identifying what tools are needed and how to build them, and developing tools for creating, structuring, describing, using, and interacting with their data and information.

To meet these HCI&IM agency needs, R&D is needed in the following areas:

- Human perceptual, cognitive, and neural processes for obtaining and using information
- The processes that computing systems use to analyze data and information, such as the uses of machine-created information
- The processes humans use (for example, observing, reading, writing, taking notes, and interacting) to obtain meaning from information
- Scientific theories of information content that allow its presentation in multiple forms and formats without changing its value
- Usability, that is, ways to provide information that is easy to create, organize, access, and use
- System use by multiple communities or user groups in multiple domains, in order to establish general methods for building capabilities that transfer across communities and domains
- Scientific theories underlying the use of models to generate data
- Developing and evaluating scientific, economic, and social domain models underlying information use

- The integration of cognitive principles of use in the specification, design, and implementation of interactive presentations, such as visualizations, so that they are appropriate for the information content
- Interoperability of data, information, and associated software
- Methods for designing, building, and maintaining systems for a usable, extendable IM environment
- Methods for assessing the effectiveness of IM systems

2.2 Managing Information as an Asset

On-line data and information provide both opportunities and challenges that require new information management methods.

The greater complexity and richness of on-line materials provides opportunities. Of our five senses, we most often use our eyes and ears, but are increasingly using touch, to interact with on-line content. For each sense, content can be presented in multiple ways. Figures 1, 2, and 3 shown above of bits, code, and maps of ocean surface speed are illustrative. The HCI field refers to sensory-based presentations as *modalities* and uses the word *mode* to denote different ways of presentation within a modality.

A movie is multi-modal since it involves both sight and sound; text and images are different modes of visual presentation. We need to develop ways to index, organize, and manage multi-modal content to be able to find it, provide it as requested, and allow user interaction with it.

Preservation provides a new challenge. Hardware platforms and software to read data on those platforms rapidly become obsolete. While we can still read 500-year-old books today, how will we access today's on-line information ten years, let alone 500 years, from now?

Some information management issues lie outside the scope of HCI&IM R&D. One is intellectual property rights, which may require legislative or regulatory action. Another is computing and networking R&D, though those fields are strongly interdependent with HCI&IM.

HCI&IM R&D per se has its own rich set of topics for R&D in managing very large, distributed, heterogeneous multi-modal collections of data and information. These topics include:

- Digitizing legacy information and creating on-line descriptions of off-line material
- Cataloging, searching, finding, discovering, viewing, processing, and disseminating data and information
- Metadata and new ways to index and find information
- Multi-modal and multi-mode access
- Interoperability of data, information, and the software that accesses and uses them
- Guaranteeing 24/7 accessibility and dissemination
- Provenance, access and version control, accuracy and integrity

- Technologies for guaranteeing security, privacy, and confidentiality
- Long-term archival storage and preservation
- Designing, building, and maintaining information management systems
- Scalability of the collections and their management

2.3 Human-Computer Interaction and Interaction Devices

Humans use computing systems to augment their own capabilities. To do so, they interact with the data and information in those computing systems, in several ways:

- They seek content from information sources.
 - The content can be presented in different modalities and different modes – visually (text, images), audibly (spoken and non-speech sounds such as music), or haptically (touch and pressure), etc.
- They interact with what they find.
 - The means they currently use include writing or drawing, speaking or singing, pointing or touching, and moving hands or eyes.
 - They use those means to further query the information, have it presented in different ways, or give directions (for example, stop a simulation, change some numbers, and restart the simulation using those new numbers).
- They control computing-enabled devices.

HCI&IM R&D investigates interaction capabilities to allow humans to use large bodies of different types of information in better-controlled ways, and develops new means, such as new devices, for interacting with computing systems. Today's HCI&IM R&D focuses chiefly on visual and audio combinations of presentation and interaction. Interacting with systems using either written or spoken natural language processing has been a long-standing goal and HCI&IM R&D area. For example, we want to be able to ask a computer a question, either spoken or written, and have it answer that question. This will require substantial advances beyond today's technologies, in which we give key words to a search engine and usually get long lists of links in return. All modalities and modes of interaction need further exploration to maximize a computing system's response to human action or to prompt a human to act. HCI&IM R&D needs include:

- Basic understanding of the internal human perceptual, cognitive, and neural processes of obtaining and using information
- Basic understanding and best employment of media, modalities, and modes to maximize human ability to seek, access, and use information
- The science of usability, that is, of delivering information in a usable manner, which includes providing optimal interaction capabilities for all possible human use
- Basic understanding of how humans share information and the methods they use
- Basic understanding of how groups of people work in a shared information space environment

- Basic understanding of how teams (groups organized to work together) work in a shared information space environment and how they evolved to become a team
- Basic understanding of how humans use information in individual and group or team problem solving, planning, decision-making, and explaining
- Understanding how computing systems use data and information to maximize synergistic human-machine capability
- Models of humans, computing systems, and the synergies between them to aid in interactive system design

In their workshop presentations, mission agencies focused on the interaction needs of their research communities. To that end, they are conducting R&D in interactive technologies including:

- Decision support
- Designing interfaces for specific tasks and for multi-tasking
- Integrating user intentions into system or interface design
- Intelligent assistive devices and technologies ranging from handheld devices to robotic assistants
- Interactions in multimodal and multimode environments
- Modeling presentation, use, and sharing of data and information
- Multimedia technologies
- Pervasive and immersive environments
- Security
- Spoken and written languages, including translation and speech to text
- Understanding of collaboration and development of collaboration technologies
- Universal accessibility
- Usability studies

2.4 Evaluation Methods and Metrics

The three HCI&IM areas discussed immediately above all require evaluation to determine whether information technologies are successful and to suggest ways for overcoming any deficiencies that are found. For both HCI and IM, there is a limited theoretical foundation. For example, only a small body of research results exists on how humans interact with information, such as what information humans use and how they use it to meet their needs. Rather, evaluation methods applied to date have been adapted from scientific approaches in other disciplines.

Much of the evaluation methodology within the HCI community is based on experimentation and the statistical analysis techniques employed by the psychological, social, and linguistic sciences. Using these methods, measures are obtained to quantify total (human, computing system, and their interaction) system performance such as time to process or quality of results. One overarching goal of this work is to develop models of cognitive processes that, in combination with models of total systems, can predict the properties of an operational design before any

implementation begins. Today, a design can be evaluated directly only in the deployed experimental state.

IM evaluation methods have been taken from the information retrieval community. These include methods for assessing the relevance of the information returned in response to queries and methods for assessing technologies that summarize the content of retrieved material. A metric used for the latter is similarity between the content of summaries generated by computing systems and those generated by humans. Such metrics have been used in the DARPA/NIST Text Retrieval Conference (TREC) evaluations. Similarly, scientifically derived assessment tools are needed for other information management areas.

Methods for using information in today's computing systems include artificial intelligence approaches to formal logical, Bayesian or other statistical methods of inference, and neural networks and genetic algorithms for classification and categorization. One defect of logical approaches is that while they can produce logically valid inferences, the "information" within those inferences may be false. Before incorporating information produced by a computing system into an on-line data or information source, evaluation criteria must be applied to determine whether or not that information is correct. In the physical sciences such as biology, chemistry, and physics, concrete measures are used to validate information. In abstract worlds outside the physical realm, human uses determine if information is valid, and in this abstract world of information, evaluation challenges abound, as seen in the following research needs:

- Theory of evaluation for IM
- Theory-based evaluation methods for HCI&IM
- Theory of evaluation for interactive information systems
- Theory of information validation
- Metrics of total system performance
 - Under basic and a variety of other conditions
 - Error bounding criteria for system acceptability and usability
- Predictive methods for human performance while using computing systems
- Models of total system function

Methods need to be developed to validate:

- Models of natural systems:
 - Physical systems such as weather
 - Biological systems such as DNA and cells
 - Cognitive systems and performance
 - Groups and teams
- Mathematical models of analytical methods
- Models of properties of information content such as uncertainty and error propagation during inference
- Models for evaluating design before implementation, especially where the human is a critical link in system use

This page intentionally left blank.

3. HCI&IM CG Government Workshop (Held October 22, 2001)

The HCI&IM CG held an all-day multi-agency workshop on October 22, 2001, to assess the participating agencies' current HCI&IM programs and their HCI&IM needs that are critical to achieving their mission objectives. This chapter summarizes each agency's presentation.

Representatives from each of the ten Federal agencies that participate in the HCI&IM CG – AHRQ, DARPA, DOE/SC, EPA, GSA, NASA, NIH, NIST, NOAA, and NSF (full names are given in Appendix 2) – provided an overview of agency programs, plans, and needs in HCI&IM R&D. These presentations were followed by initial discussions at the workshop, and in more extensive discussions in subsequent CG meetings, with the goals of documenting shared research interests (chapter 2) and coordinating R&D in these areas.

The following summaries give the affiliations of the presenters as of October 2001. A list of participating agency representatives and report contributors including their contact information is provided in Appendix 1.

3.1 DARPA

Jean Scholtz, Program Manager, Information Technology Office (ITO), Defense Advanced Research Projects Agency (DARPA), and HCI&IM CG Co-Chair, opened the meeting by welcoming the participants, thanking them for preparing their briefing materials, and outlining the workshop's objectives.

She then discussed DARPA's HCI&IM investments. DARPA is focused on developing applications that are relevant to the military. DARPA/ITO's HCI&IM efforts are chiefly in HCI, particularly programs in spoken languages (acronyms are defined in Appendix 2):

- Communicator: dialog management to enable people to converse with a computer to create, access, and manage information and solve problems
- SPINE: speech in noisy environments
- TIDES: translation, topic detection, extraction, summarization
- EARS: speech to text, including summarization and metadata extraction
- DAML: DARPA Agent Markup Language that facilitates the Semantic Web
- Augmented cognition: use fundamental understanding of cognition and guidelines for user interfaces to enhance human abilities in diverse, stressful, operational environments

Many of these efforts include extensive evaluation competitions in which solutions are tested on common data sets and performance is compared; a number of these are conducted by NIST.

Suggested new areas of HCI R&D include:

- Implicit interactions, perceptive user interfaces, and intelligent user interfaces that have autonomy to aid users in multimodal environments
- Interactions with distributed, embedded computers or with teams of robots

DARPA/ITO IM programs include:

- Bio-surveillance, focusing on receiving and correlating data from heterogeneous data sources and detecting patterns in them
- Rapid knowledge formation to encode a broad range of knowledge into reusable databases, then automatically reason across domains

Suggested areas for new IM R&D are:

- Distributed, replicable storage to support mobile uses
- Information management for large-scale distributed systems
- Context aware information retrieval
- Locally optimized information retrieval for bandwidth-constrained situations
- Abstractions of data for storage
- Secure distributed storage with automatic replication

Scholtz concluded by reiterating the need for basic scientific knowledge and understanding that can provide a firm foundation for developing technologies that effectively deliver complex functionality to users in order to meet their needs.

3.2 EPA

Steven Fine, Program Manager, Networking and Information Technology R&D, Environmental Protection Agency (EPA), presented an overview of EPA's use of information to protect human health and the environment. Such information is used to measure and predict environmental impact and to develop recommendations and regulations. Many of the challenges facing EPA today involve pollutants that move across air, water, and soil and require integrating information from different sources and disciplines. EPA is an applications-oriented agency that works closely with other agencies with related mission interests and that uses research results and technologies developed by other agencies.

Current EPA activities include:

- Multimedia (air, water, soil) Integrated Modeling System (MIMS)
- Software infrastructure and tools for constructing, composing, executing, and evaluating multimedia models

- Enablers that let environmental decision makers, scientists, and modelers work at higher conceptual levels

Planned EPA HCI&IM activities include:

- General model of environmental data and associated programming libraries
- Interdisciplinary data analysis and visualization tools

Fine identified some of EPA's HCI&IM needs:

- Integrated data analysis approaches to exploring relationships among diverse types of data, including data mining and pattern recognition technologies
- Integrated search and retrieval across multiple digital libraries
- Efficient management and distribution of large data sets
- Approaches for efficiently and appropriately resolving spatial and temporal scale differences
- Paradigms and tools for automatically building consistent systems of models and data or for identifying semantic inconsistencies in such systems
- Approaches for joint cognitive decision making for issues that involve physical, chemical, biological, social, economic, and regulatory influences and effects
- Long-term archival of large data sets

3.3 DOE/SC

Mary Anne Scott, Program Manager, Office of Advanced Scientific Computing Research (OASCR), Department of Energy/Office of Science (DOE/SC), described OASCR as seeking to discover, develop, and deploy the computational and networking tools that enable researchers in the DOE/SC scientific disciplines to analyze, model, simulate, and predict complex physical, chemical, and biological phenomena. DOE/SC funds research in advanced scientific computing, applied mathematics, computer science, and networking, and operates supercomputers, a high-performance network, and facilities such as national laboratories.

One technical challenge faced by DOE/SC is the distributed environment it manages, including distributed resources and personnel and distribution over time zones. Motivated by the changing nature of science that is increasingly multidisciplinary, compute- and data-intensive, and dependent on sophisticated instrumentation, the DOE/SC response to this challenge includes:

- A coordinated program of technology R&D and pilot projects to pioneer new approaches to flexible, secure, coordinated resource sharing through pilot collaboratories, middleware, and network research

- Scientific Discovery through Advanced Computing (SciDAC), DOE/SC's new integrated program to:
 - Create a new generation of scientific simulation codes
 - Create the mathematical and computer systems software needed for these simulation codes to effectively and efficiently use terascale (trillions of operations per second) computing systems
 - Create a collaborative software environment to enable geographically separated scientists to work together effectively as a team and access both facilities and data remotely, sometimes synchronously and sometimes not

DOE/SC's fusion energy, particle physics, and Earth system grids face problems including:

- Moving large amounts of data
- Supporting a distributed research science base
- Discovering data and information
- Providing access to experimental devices

The vision for DOE/SC's science grid is to provide:

- An environment for scientific applications
- Middleware to put together grid services to let applications talk at various levels such as at the network level
- Access to computing, visualization, and storage resources
- Public key infrastructure and certificate authority with policies that allow a community to interact with other communities both inside DOE and in other agencies and internationally
- Reliable and secure group communications

3.4 NOAA

William T. Turnbull, Director, High Performance Computing and Communications (HPCC) Office, National Oceanic and Atmospheric Administration (NOAA), began by describing NOAA's ITR&D goals, which are to:

- Accelerate the use of high-end computing technologies for weather, climate, and fisheries modeling
- Use advanced communications to provide information more rapidly and reliably
- Adapt advanced information technologies to enhance scientific productivity

Focusing on the latter, he identified the following areas of current research interest:

- Developing and using advanced visualization technologies to increase understanding of NOAA research

- Providing NOAA researchers with seamless technologies to collaborate with colleagues in NOAA, the university community, and around the world
- Developing methods of cataloging, searching, viewing, and retrieving NOAA data distributed across the Web
- Using Java to provide data in a meaningful form to forecast offices across the Nation
- Using advanced wireless and information technologies to provide near-real-time support to disaster planning, response, recovery, and mitigation

It is in crisis response that it all “comes together,” with:

- Almost real time mining of very large data sets
- Wireless communications — for example, for people on a beach after an oil spill:
 - Letting them know what the beach looked like before the spill
 - Their detailed responses about what they are seeing communicated to geographically scattered experts

Turnbull noted that whereas bandwidth was so low a few years ago that they couldn’t realize this vision, technology advances, especially in wireless networking, suggest that it’s time to try again.

NOAA’s HCI&IM gaps include:

- Managing very large environmental databases and information sets (including event recognition and sharing vocabulary and thesaurus)
- Adaptive environments
- Language integration — being able to use language interfaces and thesaurus

In discussion it was noted that NOAA shares IM issues with EPA and DOE/SC, including the need to integrate data across many different scales of time and space. NOAA often needs to do this in a matter of hours.

3.5 NIH

James C. Cassatt, Director, Division of Cell Biology and Biophysics, National Institute of General Medical Sciences (NIGMS), National Institutes of Health (NIH), began with a brief description of NIH’s organization into some two dozen institutes and its processes, in order to provide a context for the agency’s HCI&IM programs. NIH programs are both extramural (for example, funding universities and non-profit institutions) and intramural (for example, funding research at NIH facilities). NIH funds basic and applied research and develops infrastructure to support NIH research. Examples of infrastructure support are the National Cancer Institute’s Supercomputer Center and the Center for Information Technology’s Beowulf cluster.

Cassatt described the Biomedical Information Science and Technology Initiative (BISTI).⁶ Because of the biomedical community's increased use of computers to manage and analyze data and to model biological processes, the BISTI Working Group was established in 1999 to identify NIH-supported investigators' computing resource needs in the areas of hardware, software, networking, algorithms, and training, and to identify impediments biologists face in utilizing high-end computing. BISTI recommendations include:

- Establishing national programs of excellence in biomedical computing
- Establishing a new program directed toward the principles and practices of information storage, curation, analysis, and retrieval (ISCAR)
- Providing additional resources and incentives to support those who are inventing, refining, and applying the tools of biomedical computing
- Fostering a scalable national computer infrastructure, to increase computing capacity (both local and remote), in order to establish a balanced infrastructure for all computational needs

For 50 years biologists have tried to take things apart (such as a cell into its smallest components), understand what each component does, and put the resulting information back together to understand what a cell does up to what a whole organism does. This requires not just biology, but also mathematics and computer science. In addition, it requires:

- New modeling techniques
- New information and new ways to represent that information
- Imaging technologies and new ways to represent images
- Ontologies to help understand how parts work together
- Capabilities for working with large data sets
- New ways to conduct research, such as collaboration

Research within Cassatt's own institute, NIGMS, spans all areas of biology and medicine, including genetics, developmental biology, physiology, pharmacology, bio-related chemistry, cell biology, and biophysics. A new Center for Bioinformatics and Computational Biology (CBCB) has been created within NIGMS. Its research and research support includes:

- Developing and managing computational biology programs
- Generating mathematical models of biological networks
- Developing modeling and simulation tools
- Basic theoretical studies related to network organization and dynamic processes
- Developing methods for analyzing and disseminating computational models
- Funding predoctoral training and postdoctoral fellowships

⁶ <http://www.nih.gov/about/director/060399.htm>

NIH's HCI&IM research needs are similar to those of EPA, DOE/SC, and NOAA:

- Support collaboration by distributed research communities
- Develop and share large data sets and information
- Develop data and information collection, manipulation, analysis, distribution, retrieval, visualization, maintenance, and preservation methods and tools

3.6 NSF

Gary W. Strong, Program Director, Experimental and Integrative Activities Division, Computer and Information Science and Engineering (CISE) Directorate, National Science Foundation (NSF), and the HCI&IM Co-Chair at the time of the workshop, focused his presentation on the expanded efforts in HCI&IM due to the creation and growth of NSF's Information Technology Research (ITR) program. He showed an organization chart of the NSF science directorates to illustrate the breadth of the research that shares in ITR funding. NSF funds peer-reviewed proposals submitted by university-based researchers. Strong described several grants that illustrate NSF/ITR's HCI&IM investments:

- Multilingual access to large spoken archives
- Molecular ocean data assimilation
- An ensemble approach to data assimilation in the Earth sciences
- Cognitive and social design of robotic assistants
- Adaptable voice translation for minority languages
- Research on the perceptual aspects of locomotion interfaces
- Digital clay for shape input and display
- Capturing, coordinating, and remembering human experience
- Computational tools for modeling, visualizing, and analyzing historic and archaeological sites

Strong then noted some features of NSF's HCI&IM investments:

- ITR is a generic program that is highly competitive.
- Peer-reviewed projects tend toward being incremental rather than revolutionary.
- A new BioInfo program will help address computer science/bioinformatics needs.
- Informatics and data management in massive databases are still gaps since science research data sets are generally not large.

There was a discussion of NSF's possible investment in R&D in webs of sensors and associated data management for advanced warnings and for modeling the prevalence of diseases in animals, for example. Such technologies are of interest to public health organizations, and they could be applied in other areas such as weather sensors for NOAA use and chemical sensors.

Strong interacts with the intelligence community and discussed some of their needs:

- There is substantial technology for knowledge discovery in databases, but it doesn't scale sufficiently.
- There is a need for knowledge representation technologies not just for heterogeneous data but multimodal (visual, audio, haptic) data, and a need to relate information to gain understanding.
- There is a need for knowledge sharing across traditional boundaries such as sharing data across research disciplines.

These research needs illustrate how NSF-funded HCI&IM research addresses or can address other agency needs and how other agencies might leverage NSF research results.

3.7 NASA

Patricia M. Jones, Deputy Chief, Human Factors Research and Technology Division, National Aeronautics and Space Administration (NASA) Ames Research Center and Michael G. Shafto, Program Manager, Human-Centered Computing, NASA Ames Research Center, presented current NASA efforts, including:

- Enhancing distributed teamwork across a wide range of spatial and temporal scales
- Improving individual decision making through the design of integrated displays and procedures
- Frameworks for cost-effective rapid prototyping during system life-cycle management
- Design of interface content based on a task or procedure view of use

The NASA HCI&IM research focuses on technologies that enhance cognitive and collaborative performance of expert teams. NASA mission teams include users at all levels of operations:

- Mission control
- In-space scientific experimentation
- Scientific use of collected data
- Systems design, testing, evaluation, and maintenance
- Training for all participants in mission planning and rehearsal

NASA scientific domains include:

- Earth science
- Ecology
- Environmental science
- Large-scale knowledge representation and discovery
- Software engineering support and risk management for science, engineering, and operations

- Space flight control
- Space science
- Visualization
- Weather

Cross-agency collaborations include Earth science with NOAA and human-system modeling with DoD. NASA would welcome collaborations with other agencies.

NASA focuses on modeling knowledge capture and use in the context of a user's work. A critical aspect is dynamic knowledge representation:

- Generating representations to support real-time content
- Modeling visually presented knowledge and the sharing of that knowledge
- Handling multi-resolution models and displays of information
- Measuring the validity and shelf-life of information that can include uncertainty and time dependencies

Mixed-initiative systems for future NASA missions will require research on:

- Multi-tasking models to facilitate interface design
- Intelligent assistive technologies
- Interaction within multimodal environments

In the future NASA will:

- Have smaller greatly distributed teams
- Have more complex missions
- Rely more on distributed automation that operates across tremendous space and time scales (for example from very short times to multi-year)
- Need to design and plan supporting systems with confidence (including the associated operational, cost, and risk factors)

The speakers identified future directions and research gaps:

- Knowledge management across the project lifecycle for use in NASA's technology maturation model for transitioning from basic research to deployable products
- Collaborative use of information while safeguarding confidentiality, for use in risk management and safety reporting systems
- Modeling of organizations and the impact of IT on their structure and function
- Computational modeling of human performance (factors include fatigue, visual perception, and cognitive tasks), a plug-and-play architecture to tie them together, and expanding this to groups
- Advanced interfaces in zero- and micro-gravity environments

3.8 NIST

Martin Herman, Chief, Information Access Division (IAD), Information Technology Laboratory, National Institute of Standards and Technology (NIST), described NIST's mission as strengthening the U.S. economy and improving the quality of life by developing and applying measurements, standards, and technology. NIST works closely with industry, Government, and academia. Herman's Division has four programs that involve HCI&IM R&D:

- Human language technology (HLT) – Much of this work is with DARPA, NSA, and other agencies. NIST's chief HLT contribution is providing and managing evaluation infrastructures such as tasks, test collections, test protocols, scoring procedures, analysis of results, and workshops.
 - The Text Retrieval Conference (TREC), held since 1992, has as its goal to improve information search and retrieval by, for example, reducing the number of irrelevant documents found in today's searches, placing less of a burden on the user in those searches, searching in other languages, and searching for non-text material. TREC provides the premier workshop/evaluation series in information retrieval. Much of the other NIST work described below builds on the successful TREC model.
 - The TIDES program is joint with DARPA. NIST/IAD holds a Document Understanding Conference to evaluate summarization technologies, does topic detection and tracking to evaluate event detection, evaluates cross-language retrieval through TREC, and evaluates automatic content extraction.
 - Helping to advance the state of the art in speech recognition and understanding through the development of measurement methods, providing reference materials, coordinating benchmark tests, and building prototype testbed systems. Projects include automatic meeting transcription, speaker recognition, and conversational telephone speech recognition.
- Multimedia technologies – to use metrics, standards, and testing to advance technologies for accessing and using multimedia information (for example, testing multimedia searching and filtering solutions to determine their performance levels), and to advance multimedia standards through leadership on MPEG, JPEG, and Web3D standards committees (such as providing measurements and evaluation for the MPEG-7 metadata standard)
- Interaction among users and information – to provide metrics, standards, and test methods to improve the usability of interactive systems. Examples include:
 - Web Metrics, a project to develop methods for rapid remote automated usability testing of Web sites and to facilitate their integration into commercial Web development and usability testing tools through standards and proofs of concept
 - Industry USability Reporting (IUSR) to increase the visibility of software usability

- Common Industry Format: Testing, Evaluation, and Reporting (CIFter) to create benchmark test data for Web usability evaluation methods
- Providing technical guidance to GSA about testing and conformance for accessibility under Section 508 of the Americans with Disabilities Act
- Pervasive computing
 - Technologies include distributed perceptual interfaces, smart work spaces, pervasive software, security, pervasive devices, information access, dynamic service/device discovery, and wireless networking.
 - NIST's goal is to foster the adoption of pervasive (present throughout the environment) computing by providing industry, users, standards organizations, and academia with tools and tests to:
 - ◆ Identify and define product requirements
 - ◆ Develop, analyze, and compare technical solutions
 - ◆ Measure components and systems
 - ◆ Develop high quality, correct, robust implementations

It does so by providing specific methods, metrics, tests, reference data, and technologies. Benefits can include conformance and interoperability.
 - The goal of NIST's Smart Space Testbed is to apply metrics, standards, and evaluation methods to advance technologies in multimodal perceptual user interface technologies (speech, natural language, vision, etc.), multi-sensor integration, integrated information retrieval technologies, integrated mobile devices and sensors, integrated network services, and collaboration technologies

Herman identified the following gaps in HCI&IM R&D at NIST:

- Greater support in IT accessibility
- Evaluation and standards for intelligent information systems
- Immersive, interactive environments
- Universal access

3.9 GSA

Susan B. Turnbull, Director, Center for IT Accommodation, Office of Governmentwide Policy, General Services Administration (GSA), presented an overview of HCI&IM needs of GSA and the Federal CIO Council. Both are governmentwide service organizations, not science organizations. The Federal government has large buying power, reflected in its \$45 billion investment in IT products and services, which it needs to invest wisely. The Government needs to buy commercial technologies, and Turnbull wants HCI&IM R&D insights to be reflected in what is commercially available.

The goal is a citizen-connected Government that provides universal access to both science and service communities. One way is to eliminate stove-piping. Toward that end, Turnbull identified common elements of current and future plans:

- Open, standards-based Federal enterprise architecture
- Knowledge sharing among communities of practice such as scientific user communities
- Adoption of inclusive, universally accessible IT infrastructures
- Validation of IT usability and accessibility performance

Current plans include:

- Federal Bridge Certificate Authority for PKI
- Consolidation of citizen services – Web portals such as firstgov that transcend Government boundaries
- Study of public use of geographic location to find services
- ebXML to support business data reuse and integration of access channels
- Implementation of Federal enterprise architecture for e-grants
- Standardized wireless networks
- E-records management (the National Archives and Records Administration is the lead agency)

Turnbull identified gaps in HCI&IM R&D that will affect success:

- Cultural change process to improve adoption of disruptive technologies – even collaboration tools can be disruptive
- Telecommuting in relation to risks and benefits of broadband connections
- Market demand for uninterrupted access levels – providing information about new products or new versions to employees so they are not idled for long times and are not excluded from purchasing new technologies
- Methods and metrics for validating IT usability and accessibility performance, including use of wireless networking

Missing research topics include:

- The role of free or open source software
- Governance and transformation of social institutions
- Socio-technical design methods to facilitate use of IT in a mediating role
- Community-building metrics applied to IT performance
- Pragmatics of social communication to experience human-problem domain interaction and not human-machine interaction
- Speed efficiencies developed under the Federal Next Generation Internet (NGI) program, now broadened to include quality of service (QoS) effectiveness measured by success in real-life situations
- Market demand as a stronger driver than technology push of research results

- Civic science – how to tackle complex problems (such as economic and health problems) at the source, in light of global interdependence

3.10 AHRQ

J. Michael Fitzmaurice, Fellow of the American College of Medical Informatics (FACMI), Senior Science Advisor for Information Technology, Immediate Office of the Director, Agency for Healthcare Research and Quality (AHRQ), Department of Health and Human Services (DHHS), described AHRQ as a research agency that studies how to improve the effectiveness of health care services and the health care delivery system. Congress has directed the agency to:

- Identify the causes of preventable health care errors and patient injury in health care delivery
- Develop, demonstrate, and evaluate strategies for reducing errors and improving patient safety
- Disseminate such effective strategies throughout the health care industry

AHRQ works to improve decisions in three areas – clinical, health care, and public policy. The agency funds scientific research and synthesizes and develops evidence-based medical information to inform the health care choices of health care practitioners and patients. AHRQ research advances the use of information technology for coordinating patient care and for conducting quality and outcomes research. AHRQ goals are to understand patient safety incidents and reduce medical errors and injuries through the use of information technologies, such as the Internet, that enhance information access, provider-patient communication, and shared decision-making. AHRQ research programs focus on IT in two areas:

- Improving patient safety
- Improving the quality of patient care by evaluating clinical applications that use health data standards and information technology

An example is the development and testing of handheld and other portable devices by patients and providers to access critical information or remotely monitor patients.

AHRQ supports research to develop and validate quality measures and to provide access to clinical practice guidelines to better inform decision making by both providers and patients. The guidelines are in two clearinghouses:

- The National Guidelines Clearinghouse, an Internet-based repository of about 1,000 clinical practice guidelines or links to such guidelines
 - AHRQ partners are the American Medical Association and the American Association of Health Plans
- The Internet-based National Quality Measures Clearinghouse contains detailed information about quality measures (such as appropriateness, timeliness, whether the

desired care outcome was achieved, and satisfaction), their validity, and how to access them

AHRQ is developing an umbrella Web site to link these two clearinghouses through common medical terminology. An umbrella of quality tools, including this site, will use standard Web-based data query, database storage software, and search engine tools, and will have consistent graphic and navigation features and controlled vocabularies.

AHRQ also funds research that supports the States as well as the Federal government in patient safety incident data reporting, analyzing causes, providing feedback, and remedies such as education (for example, computerized prompts, alerts, and reminders) and training.

Four DHHS agencies are working to integrate Federal medical event reporting input and reporting data structure such as an integrated Web-based reporting interface. The four agencies are FDA (for adverse drug reactions), CDC (communicable diseases), CMS (quality of care), and AHRQ. Their overall goal is to have a more comprehensive system that covers incidents such as “no harm events” and “near misses” in addition to deaths due to errors addressed in the Institute of Medicine report *To Err is Human*.⁷ Detailed goals include:

- Ease of use
- Reduced reporting burden
- Providing reliable and valid information to Federal, State, and private users
- Maintaining confidentiality
- Reflecting and supporting public-private collaboration
- Most important, resulting in improvement

⁷ National Academies Press, Washington, D.C., 2000.

4. OTHER GOVERNMENT PROGRAMS

The HCI&IM CG sought input about other Government programs that are working on solutions to HCI&IM problems and on ways their work can contribute to integrated HCI&IM R&D. Representatives of three such programs described their work at later HCI&IM CG meetings.

4.1 NSF's Digital Libraries and Digital Government Programs and their Joint Workshop with the Library of Congress on Preservation

On October 28, 2002, Lawrence E. Brandt, Digital Government (DG) Program Manager at NSF, and Stephen M. Griffin, Program Manager for the Digital Libraries (DL) Program at NSF gave a three-part briefing that summarized their individual programs and then focused on a workshop they co-sponsored with the Library of Congress.

- The DG Program supports multi-disciplinary cross-sector (academia, industry, and Government) collaborative information technology research to improve the availability and use of on-line government information.
- DL research is grounded in computer science and engineering research, informed by domain research across disciplines, applicable to a broad set of scientific and non-scientific problem domains and characterized by novel collaborative efforts focused on the creation, collection, organization, use, and preservation of large volumes of digital information in a rapidly changing, globally-linked knowledge environment.
- The DG and DL programs and the Library of Congress jointly sponsored a Workshop on Research Challenges in Digital Archiving that was held at the Library of Congress on April 12-13, 2002. The workshop's goal was to better understand issues in the long-term preservation of emerging vast on-line resources. The issues are of immediate concern due to the ephemeral nature of on-line material and the speed at which the underlying technology changes, leading to increased loss over time. While preserving bits is necessary and challenging, logical preservation is also needed. Challenges particular to digital archiving include the long-term perspective; interrupted management; changing user communities, requirements, and expectations; and systems and technologies evolving around the data. Research priorities include:
 - Reference architecture that supports research
 - Metrics for cost, value, policy options, outcomes
 - Preservation of dynamic objects
 - Decision models
 - Predictive models of use and value
 - Scalability up and down
 - Tools for ingest, metadata, and push/pull

- Incentives for building persistent archives and for depositors
- User requirements

Other issues that need to be addressed include economic or business models, intellectual property rights, and policies. The workshop considered various technical frameworks as good starting points.

4.2 Air Force Research Laboratory Human Effectiveness Directorate

Illustrating HCI&IM research by the DoD services, representatives from three programs in the Human Effectiveness Directorate at the Air Force Research Laboratory (AFRL), Wright-Patterson Air Force Base, Dayton, Ohio, briefed the HCI&IM CG on their research programs on February 24, 2003.

Maris M. Vikmanis, Chief of the Crew System Interface Division, began by describing his Division's mission, which is to conceive, develop, integrate, and transition information display and performance aiding technology to assure the preeminence of U.S. air and space forces. The Division's branches address aspects of the mission:

- Visual Display Systems Branch – helmet displays, flat panel displays, night vision systems, sensor cueing, vision science, optics/transparencies
- Aural Displays and Bioacoustics Branch – voice communications, acoustic signatures, vibration, 3-D audio, noise suppression
- Human Interface Technology Branch – multi-sensor displays, hands-free control, crew station fit, operator functional state, aiding and adaptation
- Information Analysis and Exploitation Branch – speech processing, voice recognition, battlespace awareness, cognitive display
- Crew System Development Branch – requirements, analysis, design integration, test and evaluation, cockpit concepts, modeling and simulation tools, real-time mission simulation

Vikmanis identified two growth areas:

- Decision-centered command and control, including software interface agents, information visualization, and cognitive task analysis
- Remotely controlled vehicles, including visual interface requirements for unmanned air vehicles (UAV) and satellite control, supervision of highly automated systems, and cost-effective telepresence

Robert Eggleston, Senior Scientist, described R&D on a Work-Centered Support System (WCSS). The goal of its interface client technology is to provide a full spectrum of aids to individuals, work cells, and distributed work groups that support flexible and adaptive work practices to handle, for example, interruptions, emergent procedures, work arounds, and time pressure.

Two examples are:

- Flight management support: To address cognitive challenges such as juggling multiple on-going flight plans, getting current data and meeting multiple constraints, detecting and responding to conflicts, and handling constant interruptions, a new “ecological” user interface is being developed. This interface is organized around how users work, reduces cognitive/perceptual burdens, and is resistant to interruptions.
- Weather forecasting in support of flight operations, where an interface is being developed to address the primary cognitive challenge of integrating high-density data from multiple sources and in multiple forms, keeping track of changing events and determining their significance, and handling ambiguous and conflicting data.

Kathleen M. Robinette, Principal Research Anthropologist, described the World Engineering Anthropometry Resource (WEAR), consisting of a 3-D database of scanned images of more than 4,000 people and associated analysis tools and user interface. The database reflects NATO’s age, gender, and ethnicity variability. It is being used in designs ranging from cockpits to apparel. Challenges include:

- 3-D data collection standards and terminology standards
- 3-D shape image cataloging and searching
- Privacy and security

4.3 DOE/OSTI and CENDI

On March 3, 2003, Walter L. Warnick, Director, DOE/SC’s Office of Scientific and Technical Information (OSTI), briefed the HCI&IM CG. He described OSTI’s mission of collecting, announcing, and making available scientific and technical information (STI) to advance scientific knowledge. While there are three billion Web pages that engines such as google can easily search, the “deep Web” contains vast but hard to search resources such as huge governmental databases. Web sites such as science.gov let users search both these databases and other Web sites that contain technical reports, journal articles, and other published material. Challenges OSTI is addressing include:

- Knowing if a search has overlooked a key resource. A solution is to take advantage of today’s network speed to search every possible source.
- Developing an architecture that collects search results to selectively allow user downloads, which is especially useful for large distributed data sets

Warnick also described the work of CENDI, an interagency working group of senior STI managers from nine U.S. agencies. The CENDI acronym collects the first letters in the names of its participating departments and agencies: Commerce; Energy and EPA; NASA and the

National Libraries of Agriculture, Education, and Medicine; Defense; and Interior. While many of these agencies have R&D programs that are part of HCI&IM, the STI managers fulfill the non-research function of providing scientific information to the scientific user community. CENDI goals are to provide coordination and leadership for information exchange on STI policy, promote the development of improved STI systems, and promote understanding of STI and STI management.

CENDI has held workshops on:

- Business Continuity and Disaster Recovery
- Managing and Preserving Electronic Resources: The OAIS Reference Model
- PKI and Digital Signatures: From E-Commerce to E-Information Management
- Evaluating Our Web Presence: Challenges, Metrics, Results
- Handles as Persistent Identifiers

CENDI projects and reports include:

- Science.gov
- Copyright FAQs
- License Agreements for Electronic Products and Services: FAQ
- Evaluating Our Web Presence: Challenges, Metrics, Results

Appendix 1: Workshop Attendees and Contributors to this Report

William S. Bainbridge, Ph.D.

*Deputy Division Director,
Information and Intelligent
Systems Division, Directorate
for Computer and Information
Science and Engineering,
National Science Foundation
4201 Wilson Boulevard
Suite 1115
Arlington, VA 22230
(703) 292-8930
FAX: (703) 292-9073*

Lawrence E. Brandt

*Program Manager,
Experimental and Integrative
Activities Division, Directorate
for Computer and Information
Science and Engineering,
National Science Foundation
4201 Wilson Boulevard
Suite 1160
Arlington, VA 22230
(703) 292-8980
FAX: (703) 292-9030*

James C. Cassatt, Ph.D.

*Director, Division of Cell
Biology and Biophysics,
National Institute of General
Medical Sciences, National
Institutes of Health
4500 Center Drive
MSC 4500
Bethesda, MD 20892
(301) 594-0828
FAX: (301) 480-2004*

Paul Domich, Ph. D.

*Associate Director, Building
and Fire Research Laboratory,
National Institute of Standards
and Technology
325 Broadway
Mail Code 809.02
Boulder, CO 80305-3328
(303) 497-5112*

*At the time of the workshop,
Domich was Senior Policy
Analyst, Office of Science and
Technology Policy*

Robert Eggleston

*Senior Scientist, Air Force
Research Laboratory
2255 H Street
Wright-Patterson AFB, OH
45433-7022*

Steven Fine, Ph.D.

*Program Manager, Networking
and Information Technology
R&D, Environmental Protection
Agency
MD E243-01, EPA
Research Triangle Park, NC
27711
(919) 541-0757
FAX: (919) 541-1379*

**J. Michael Fitzmaurice, Ph.D.,
FACMI**

*Senior Science Advisor for
Information Technology,
Immediate Office of the
Director, Agency for Healthcare
Research and Quality
540 Gaither Road, Suite 3026
Rockville, MD 20850
(301) 427-1227
FAX: (301) 427-1210*

Helen M. Gigley, Ph.D.

*Human-Computer Interaction
and Information Management
(HCI&IM) Liaison and Software
Design and Productivity (SDP)
Liaison, National Coordination
Office for Information
Technology Research and
Development
4201 Wilson Boulevard
Suite II-405
Arlington, VA 22230
(703) 292-4873
FAX: (703) 292-9097*

Stephen M. Griffin

Program Manager, Digital Libraries Initiative, Information and Intelligent Systems Division, Directorate for Computer and Information Science and Engineering, National Science Foundation
4201 Wilson Boulevard
Suite 1115
Arlington, VA 22230
(703) 292-8930
FAX: (703) 292-9073

Brian Gross

Deputy Director, Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration
Forrestal Campus,
U.S. Route 1
P.O. Box 308
Princeton, NJ 0852
(609) 452-6504
FAX: (609) 987-5063

Martin Herman, Ph.D.

Chief, Information Access Division, Information Technology Laboratory, National Institute of Standards and Technology
100 Bureau Drive, Stop 8940
Gaithersburg, MD 20899-8940
(301) 975-4495
FAX: (301) 670-0939

Sally E. Howe, Ph.D.

Associate Director, National Coordination Office for Information Technology Research and Development
4201 Wilson Boulevard
Suite II-405
Arlington, VA 22230
(703) 292-4873
FAX: (703) 292-9097

Patricia M. Jones, Ph.D.

Deputy Chief, Human Factors Research and Technology Division, NASA Ames Research Center, National Aeronautics and Space Administration
MS 262-11, Moffet Field, CA
94035-1000
(650) 604-1345
FAX: (650) 604-3323

Grant Miller, Ph.D.

Large Scale Networking (LSN) Liaison, National Coordination Office for Information Technology Research and Development
4201 Wilson Boulevard
Suite II-405
Arlington, VA 22230
(703) 292-4873
FAX: (703) 292-9097

Kathleen M. Robinette

Principal Research Anthropologist, Air Force Research Laboratory
2255 H Street
Wright-Patterson AFB, OH
45433-7022
(937) 255-8810
FAX: (937)-255-8752

Jean Scholtz, Ph.D., Co-Chair, HCI&IM

Information Access Division, Information Technology Laboratory, National Institute of Standards and Technology
100 Bureau Drive, Stop 8940
Gaithersburg, MD 20899-8940
(301) 975-2520
FAX: (301) 975-5287

At the time of the workshop, Scholtz was Program Manager, Information Technology Office, Defense Advanced Research Projects Agency

Mary Anne Scott, Ph.D.

*Program Manager,
Mathematical, Information, and
Computational Sciences
Division, Office of Advanced
Scientific Computing, Office of
Science, Department of Energy
OASCR/MICS, SC-31
1000 Independence Ave., SE
Washington, D.C. 20874-1290
(301) 903-6368
FAX: (301) 903-7774*

Michael G. Shafto, Ph.D.

*Program Manager, Human-
Centered Computing, NASA
Ames Research Center, National
Aeronautics and Space
Administration
Mail Stop 269-4
Moffet Field, CA 94035-1000
(650) 604-6170
FAX: (650) 604-6009*

Gary W. Strong, Ph.D.

*Manager, Behavioral Research
and Biometrics, Department of
Homeland Security
(202) 772-9659
FAX: (202) 772-9720*

*At the time of the workshop,
Strong was Co-Chair, HCI&IM
CG and Program Director,
Experimental and Integrative
Activities Division, Directorate
for Computer and Information
Science and Engineering,
National Science Foundation*

Susan B. Turnbull

*Senior Program Advisor,
General Services Administration
and Office of Intergovernmental
Solutions, Office of Citizen
Services and Communications
Emerging Technology
Subcommittee, Architecture and
Infrastructure Committee, CIO
Council
1800 F Street, NW (MKC)
Room 2236
Washington, D.C. 20405
(202) 501-6214
FAX: (202) 219-1533*

*At the time of the workshop,
Turnbull was Director, Center
for IT Accommodation, GSA*

William T. Turnbull

*Deputy CIO and Director, High
Performance Computing and
Communications (HPCC)
Office, National Oceanic and
Atmospheric Administration
Room 9600
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-9600
FAX: (301) 713-4040*

*At the time of the workshop,
Turnbull was Deputy Director,
HPCC Office, NOAA*

Maris M. Vikmanis

*Chief, Crew System Interface
Division, Air Force Research
Laboratory
2255 H Street
Wright-Patterson AFB, OH
45433-7022
(937) 255-7580*

Walter L. Warnick, Ph.D.

*Director, Office of Scientific and
Technical Information, Office of
Science, Department of Energy
P.O. Box 62, Oak Ridge, TN
37831
(865) 576-1189
FAX: (865) 576-3609*

This page intentionally left blank.

Appendix 2: Acronyms

AFRL – DoD’s Air Force Research Laboratory

AHRQ – DHHS’s Agency for Healthcare Research and Quality

BISTI – NIH’s Biomedical Information Science and Technology Initiative

Bit – basic binary digit in the computer

Byte – group of adjacent binary digits operated upon as a unit (usually connotes a group of eight bits)

CBCB – NIH/NIGMS’s Center for Bioinformatics and Computational Biology

CDC – DHHS’s Centers for Disease Control and Prevention

CENDI – interagency working group of senior scientific and technical information managers from nine U.S. agencies. The CENDI acronym includes the first letter of participating departments or agencies: Commerce; Energy and EPA; NASA and the National Libraries of Agriculture, Education, and Medicine; Defense; and Interior

CG – Coordinating Group

CIFter – Common Industry Format: Testing, Evaluation, and Reporting

CIO – Chief Information Officer

CMS – DHHS’s Center for Medicare and Medicaid Services

DAML – DARPA Agent Markup Language, which facilitates the Semantic Web

DARPA – DoD’s Defense Advanced Research Projects Agency

DG – Digital Government

DHHS – Department of Health and Human Services

DL – Digital Libraries

DOC – Department of Commerce

DNA – Deoxyribo Nucleic Acid

DoD – Department of Defense

DOE/SC – Department of Energy’s Office of Science

EARS – Effective, Affordable, Reusable Speech-to-Text

ebXML – electronic business XML initiative designed to standardize XML business specifications for the exchange of all electronic business data

EPA – Environmental Protection Agency

FACMI – Fellow of the American College of Medical Informatics

FAQ – frequently asked questions

FDA – DHHS’s Food and Drug Administration

GFDL – NOAA’s Geophysical Fluid Dynamics Laboratory

Gigabytes – 10^9 bytes

GSA – General Services Administration

HCI&IM – Human-computer Interaction and Information Management

HLT – human language technology

HPCC – High Performance Computing and Communications

IAD – NIST’s Information Access Division

ISCAR – information storage, curation, analysis, and retrieval

ITO –DARPA’s Information Technology Office

ITR – NSF’s Information Technology Research program

IUSR – Industry Usability Reporting

JPEG – standardized still image compression mechanism created by the Joint Photographic Experts Group

MIMS – Multimedia Integrated Modeling System

MPEG – Multimedia Content Description Interface created by the Moving Picture Experts Group

NASA – National Aeronautics and Space Administration

NCO/ITRD – National Coordination Office for Information Technology Research and Development

NGI – Federal Next Generation Internet program

NIGMS – NIH’s National Institute of General Medical Sciences

NIH – DHHS’s National Institutes of Health

NIST – DOC’s National Institute of Standards and Technology

NITRD – Networking and Information Technology Research and Development

NOAA – DOC’s National Oceanic and Atmospheric Administration

NSF – National Science Foundation

OAIS – Open Archival Information System

OASCR – DOE/SC’s Office of Advanced Scientific Computational Research

OSTI – DOE/SC’s Office of Scientific and Technical Information

PKI – Public Key Infrastructure

QoS – quality of service

R&D – research and development

SciDAC – DOE/SC’s Scientific Discovery through Advanced Computing program

SPINE – speech in noisy environments

STI – Scientific and Technical Information

Terascale – trillions of operations per second

TIDES – Translingual Information Detection, Extraction, and Summarization

TREC – Text Retrieval Conference

UAV – unmanned air vehicles

WEAR – AFRL’s World Engineering Anthropometry Resources

Web3D – programming or descriptive languages that can be used to deliver interactive 3D objects and words across the Internet

WCSS – AFRL’s Work-Centered Support System

XML – Extensible Markup Language, the universal format for structured documents and data on the Web

Acknowledgements

The HCI&IM CG Co-Chairs thank the many participants and contributors to the HCI&IM workshop and to this report. The Co-Chairs especially thank Dr. Helen M. Gigley who made major contributions to the organization of the workshop and to the content of this report while serving since 2000 as the NCO/ITR&D's liaison to the HCI&IM CG. Dr. Gigley's expertise and insights were essential to making this report possible.

Copyright

This is a work of the U.S. Government and is in the public domain. It may be freely distributed and copied, but it is requested that the National Coordination Office for Information Technology Research and Development (NCO/ITR&D) be acknowledged. Please ask the provider of the image for permission to use it.

National Coordination Office for
Information Technology Research and Development

Suite II-405
4201 Wilson Blvd.
Arlington, VA 22230
(703) 292-4873

Email address: nco@nitrd.gov

Web address: <http://www.nitrd.gov>